

Modélisation de réseaux de régulation de gènes dynamiques et prédiction d'expériences d'interventions biologiques dans des cellules cancéreuses.

R. Loubaton, N. CHAMPAGNAT, L. VALLAT, P. VALLOIS

Université de Lorraine, Université de Lorraine, Université de Strasbourg, Université de Lorraine

Email : rodolphe.loubaton@univ-lorraine.fr

Mots Clés : Mathématiques appliquées au domaine médical, cancer, réseaux de régulation de gènes, régression linéaire pénalisée (*lasso*), inférence statistique, modélisation stochastique.

Biographie – Après mes études supérieures à Agrocampus Ouest où j'ai obtenu un master en statistique appliquée et Data Science, je réalise une thèse de mathématiques appliquées au domaine médical, à l'Institut Elie Cartan de Lorraine à Nancy, encadrée par Nicolas Champagnat (DR INRIA) et par Laurent Vallat (MCUPH à Strasbourg). Ma thèse, intitulée "Caractérisation de cibles thérapeutiques dans un programme génique tumoral", est financée par la région Grand-Est.

Resumé : Les cellules de l'organisme sont en contact permanent avec leur environnement et des modifications de celui-ci induisent des réponses adaptées des cellules. Ces réponses cellulaires se font à travers l'expression de milliers de gènes qui s'activent de façon concertée et ces interactions sont représentées sous forme d'un réseau de régulation de gènes.

Dans le cas des cellules cancéreuses, il existe des anomalies de l'expression de certains gènes qui altèrent le fonctionnement normal de ces réseaux de gènes et protéines et donc des programmes géniques, qui correspondent à la mise en marche, au contrôle et à l'arrêt de l'activité d'ensembles de gènes dans le cadre de la réalisation ordonnée d'une opération biologique qui exige de nombreux intervenants moléculaires. Ces anomalies, responsables du dysfonctionnement des programmes géniques qui contrôlent la division cellulaire, provoquent une prolifération incontrôlée des cellules.

Le projet général de cette thèse, qui s'inscrit dans la continuité des travaux de Laurent Vallat et al. [2, 3], consiste à modéliser le fonctionnement de ces programmes géniques à partir de données temporelles d'expression de gènes et protéines. Un tel modèle permettra d'identifier les gènes dont la modulation d'expression permettra d'améliorer le comportement cellulaire dans un but thérapeutique.

Laurent Vallat et son équipe ont prélevé des cellules saines et cancéreuses chez des sujets contrôles et des patients. Puis ils ont mesuré, à différents instants, l'expression des gènes par RNA-seq (mesure du nombre de copies d'ARN de plus de 20 000 gènes) et l'expression des protéines par spectrométrie de masse (quantité d'environ 5 000 protéines) après une modulation de l'environnement cellulaire. Cette modulation est réalisée par stimulation d'un récepteur membranaire.

Tout d'abord, j'ai travaillé sur un code (logiciel R) qui permet réaliser une pré-analyse d'un jeu de données RNA-seq. Cette pré-analyse, basée sur le *package* R DESeq2 [1], permet notamment de sélectionner parmi l'ensemble des gènes (environ 20 000), un sous-ensemble de gènes "d'intérêt" (environ 1000 gènes).

J'ai ensuite construit un modèle statistique paramétrique gaussien qui permet de rendre compte de la dynamique d'expressions de gène à l'aide de paramètres décrivant les interactions entre ces gènes "d'intérêt". L'inférence statistique des paramètres du modèle combine des étapes de régressions linéaires pénalisées (*lasso*) et non pénalisées.

J'ai ensuite étudié la modélisation d'expériences biologiques modifiant les gènes. Dans la pratique, il s'agit majoritairement d'inhibition de l'expression de gènes (ou extinction de gènes ou *silencing*). Nous avons développé une méthode dite "mécaniste" et une seconde dite "conditionnelle" qui permettent de construire deux nouveaux modèles associés à une intervention biologique. Enfin, j'ai développé un code pour l'inférence et les deux modèles de prédiction.

Références

- [1] Michael I. Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, December 2014.
- [2] Cedric Schleiss, Raphael Carapito, Luc-Matthieu Fornecker, Leslie Muller, Nicodème Paul, Ouria Tahar, Angélique Pichot, Manuela Tavian, Alina Nicolae, Laurent Miguët, Laurent Mauvieux, Raoul Herbrecht, Sarah Cianferani, Jean-Noël Freund, Christine Carapito, Myriam Maumy-Bertrand, Seiamak Bahram, Frédéric Bertrand, and Laurent Vallat. Temporal multiomic modeling reveals a B-cell receptor proliferative program in chronic lymphocytic leukemia. *Leukemia*, 35(5):1463–1474, 2021.
- [3] Laurent Vallat, Corey A Kemper, Nicolas Jung, Myriam Maumy-Bertrand, Frédéric Bertrand, Nicolas Meyer, Arnaud Pocheville, John W Fisher, John G Gribben, and Seiamak Bahram. Reverse-engineering the genetic circuitry of a cancer cell with predicted intervention in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, 110(2):459–64, Jan 2013.